



Molecular Communication for Equilibrium State Estimation in Biochemical Processes on a Lab-on-a-Chip

Bayram Akdeniz, Malcolm Egan

► To cite this version:

Bayram Akdeniz, Malcolm Egan. Molecular Communication for Equilibrium State Estimation in Biochemical Processes on a Lab-on-a-Chip. IEEE Transactions on NanoBioscience, Institute of Electrical and Electronics Engineers, In press. hal-03151618

HAL Id: hal-03151618

<https://hal.archives-ouvertes.fr/hal-03151618>

Submitted on 24 Feb 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Molecular Communication for Equilibrium State Estimation in Biochemical Processes on a Lab-on-a-Chip

Bayram Cevdet Akdeniz and Malcolm Egan

Abstract—A basic problem in molecular biology is to estimate equilibrium states of biochemical processes. To this end, advanced spectroscopy methods have been developed in order to estimate chemical concentrations *in situ* or *in vivo*. However, such spectroscopy methods can require special conditions that do not allow direct observation of the biochemical process. A natural means of resolving this problem is to transmit chemical signals to another location within a lab-on-a-chip device; that is, employing molecular communication in order to perform spectroscopy in a different location. In this paper, we develop such a signaling strategy and estimation algorithms for equilibrium states of a biochemical process. In two biologically-inspired models, we then study via simulation the tradeoff between the rate of obtaining spectroscopy measurements and the estimation error, providing insights into requirements of spectroscopy devices for high-throughput biological assays.

I. INTRODUCTION

A key challenge in molecular biology is to understand how environmental conditions affect the dynamics of biochemical processes. For example, under what conditions does bacteria chemotaxis [1] occur? Or, in the context of quorum sensing [2], when does a bacteria colony exhibit a population response to increases in density? This type of question can often be rephrased in terms of the equilibrium state of complex chemical reaction systems—i.e., the long-term values for the concentration of each chemical species—which varies depending on the external environment (e.g., temperature or chemical composition) [3].

To answer these questions in practice, it is necessary to measure the concentration of chemical species produced by the biochemical process under observation—known as an assay. However, it is not always straightforward to perform such measurements *in situ* or *in vivo*. Nevertheless, they are required in order to understand how the environmental biochemistry affects the biochemical process.

To perform measurements of concentrations *in situ* or *in vivo*, advanced spectroscopy measurement techniques are required (see, e.g., [4], [5]). Often these techniques require special environmental conditions; for example, in force spectroscopy utilizing atomic force microscopy, well-defined, flat and non-reactive substrates are required [6]. In this case, it may be desirable to extract samples from the process and perform spectroscopy elsewhere.

On the other hand, recent advances in microfluidic technologies now provide a means to both house a biochemical process and perform spectroscopy on a single chip; often referred to as a lab-on-a-chip (LoC) or micro total analysis system [7]. To perform spectroscopy on the same chip, it can be desirable to transfer chemical signals from the biochemical process to another location on the chip, which can be viewed as a form of molecular communication (MC).

A number of previous works have investigated microfluidic systems from a MC perspective. In [8], the pulse width, pulse delay and pulse amplitude were studied under a convection-diffusion-reaction channel. In [9], [10], microfluidic circuits were developed in order to produce a desired pulse shape at the transmitter and perform detection at the receiver. In [11], [12], droplet-based microfluidic channels were characterized and the information capacity evaluated. Further work investigating the information capacity of microfluidic channels has been carried out in [13]–[17], often under the assumption of laminar flow modeled by convection-diffusion equations.

In this paper, we consider a biochemical process under observation contained in a chamber, connected to spectroscopy chamber—where the quantity of a given chemical species can be directly observed—via a microfluidic channel, illustrated in Fig. 1. Based on the observations from the spectroscopy chamber, the aim is to reliably estimate the equilibrium state of the biochemical process under investigation.

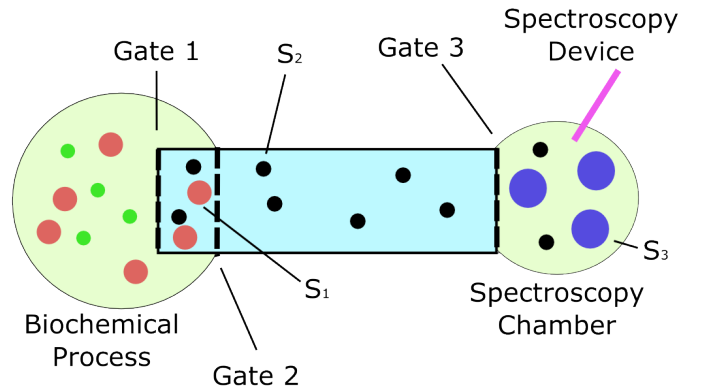


Fig. 1: System Model.

also true that the small scale of the channel can facilitate rapid diffusion [18]. For example, diffusion-based microfluidics has a long history in validating stochastic diffusion models [19]. As such, we focus on microfluidic channels with dynamics purely driven by reactions and diffusion, without convection.

The absence of convection often means that, under certain reversibility conditions on the chemical reactions in the system, that the statistics for the quantity of molecules in the spectroscopy chamber converges to an steady state. Moreover, this steady state can be readily characterized, depending only on the total quantity of molecules in the system, the reaction rates, and the volume [20]; all of which can be readily estimated.

As such, it is feasible to utilize observations from the spectroscopy chamber in order to estimate the equilibrium state of the biochemical process. There are three scenarios: (i) the set of equilibrium states is known completely, either from a mechanistic description or from previous experimentation; (ii) the set of equilibrium states is unknown; and (iii) the set of equilibrium state is partially known. Scenario (i) bears strong similarities to the standard MC framework, where it is necessary to detect one of a set of known messages. On the other hand, in Scenarios (ii) and (iii), the set of equilibrium states must be estimated.

Aside from the fact that the set of equilibrium states (or messages) is unknown in Scenarios (ii) and (iii), another key difference from standard MC schemes is that even in Scenario (i), signaling is often non-binary (corresponding to multiple equilibrium states of the biochemical process). Traditionally, MC systems have struggled with non-binary message sets due to high levels of noise unless multiple chemical species are utilized [21]. Nevertheless, in our approach, we show that multiple possible equilibrium states in Scenario (i) can be dealt with via multiple samples from a single experiment. In Scenarios (ii) and (iii), we develop an estimation procedure by exploiting a Gaussian mixture model. This is feasible due to the fact, shown in the sequel, that the steady state statistics for the quantity of molecules in the spectroscopy chamber can be well approximated via a Gaussian distribution.

We illustrate our estimation procedure on two biologically-inspired models for the biochemical process. The first model is based on a self-actuating toggle switch circuit, which plays an important role in cellular differentiation [22]. A key feature of the toggle switch circuit is that it can admit tristability; that is, the circuit can produce three distinct equilibrium states. As such, a non-binary signaling scheme is required. The second model is based on the signaling pathway involved in bacterial chemotaxis—an example of the phenomenon of robust adaptation—where the chemical species to be observed is based on the CheY protein governing rotation of the bacterial rotor [23].

In both models, we numerically investigate the performance of our estimation procedure under each scenario. In particular, we obtain tradeoffs between the number of required samples on the probability an equilibrium state is misidentified (called a clustering error), and on the mean-square error of the estimated equilibrium states in Scenarios (ii) and (iii). These results provide guidelines for the required rate of spectroscopy

measurements for high-throughput biological assays.

II. SYSTEM MODEL

A. Biochemical Assay Setup

Consider a microfluidic chip consisting of two chambers and a microchannel, as illustrated in Fig. 1. The first chamber houses the biochemical process under investigation. This chamber is separated from the microchannel by two gates. The first gate is opened once the biochemical process reaches an equilibrium state, allowing molecules S_1 from the first chamber to diffuse inside. The second gate acts as a filter, allowing only a further information-carrying molecule S_2 , distinct from any of the molecules produced by the biochemical process, to pass through. After a short period of time, the first gate is closed, allowing no further molecules from the biochemical process to enter the beginning of the microchannel. We expect that such gates can be implemented using microfiltration techniques [24].

When molecules of species S_1 from the biochemical process enter the beginning of the microchannel, they are able to react to form information-carrying molecules of species S_2 (detailed in Sec. II-B). The molecules of S_2 can then pass through the second gate and diffuse through the remainder of the microchannel towards the spectroscopy chamber.

Within the spectroscopy chamber, the molecules of S_2 can then react to form molecules of species S_3 (detailed in Sec. II-B). Between the microchannel and the spectroscopy chamber is a third gate, which prevents molecules of species S_3 to pass back into the microchannel, while allowing molecules of species S_2 .

The spectroscopy chamber is equipped with a spectroscopy device able to passively count how many molecules of species S_3 are present. The precise form of the spectroscopy device depends on the species S_3 . The spectroscopy device then passes the count to an external processor, which is used to estimate the quantity of molecules of S_1 which were present initially at the beginning of the microchannel between the first and second gates.

A concrete example of this setup can be developed for detection of autoinducer molecules arising in bacteria colonies for the purpose of quorum sensing. In this case, each bacteria forms a biochemical switch; able to produce either a small or large quantity of autoinducer molecules depending on the density of the bacteria colony.

In order to observe the quantity of autoinducer, Raman spectroscopy has been exploited in [4] for bacteria colonies *in situ*. This previous work does not consider distributed spectroscopy; i.e., the spectroscopy is performed in the same chamber as the biochemical process. Nevertheless, it is feasible that, equipped with a LoC, the spectroscopy could be performed in a separate chamber. In fact, Raman spectroscopy has been proposed as a means of cooperative in-vivo sensing in [25]. We also highlight that a distributed implementation is likely to be necessary for other systems, where the spectroscopy method may perturb the biochemical process or the detection chamber requires special preparation (such as in force spectroscopy [6]).

B. Dynamics of the Microfluidic Channel

The portion of the first chamber used to produce signaling molecules, the microchannel, and the detection chamber are assumed to form a domain $\Omega \subset \mathbb{R}^d$, $d \in \{1, 2, 3\}$ with smooth boundary $\partial\Omega$. Consider the discretization of Ω into N volume elements (voxels) each of volume V_{vox} , with the subdomain forming voxel i denoted by \mathcal{V}_i , $i = 1, \dots, N$. Here, volume is interpreted as length in \mathbb{R}^1 , area in \mathbb{R}^2 , and volume in \mathbb{R}^3 .

The portion of the first chamber housing the biochemical process used to produce the signaling molecules, called the *transmitter*, has volume V_{Tx} . The production of signaling molecules is achieved via the unimolecular reactions



In particular, the transmitter produces information-carrying molecules of species S_2 by the first reaction in (1). In general, a more complex reaction pathway may be present; however, we assume in this case that the intermediate reactions occur rapidly as is common, for example, in enzyme-based reactions [20]. We also note that unimolecular reactions are capable of modeling the dynamics of a range of biochemical systems [26].

We assume that molecules of species S_1 produced in the transmitter are not capable of diffusing into the channel. On the other hand, this is possible for species S_2 .

In the detection chamber, called the *receiver*, with volume V_{Rx} (not necessarily the same as V_{Tx}), molecules of species S_2 are able to generate molecules of species S_3 via



The molecules of S_3 are then used for the purpose of spectroscopy (e.g., Raman or force spectroscopy) in the detection chamber. We assume that the spectroscopy is passive, which means that no molecules of S_3 are removed by the detection process. As such, the detection process can be viewed as a passive receiver, which has been considered in previous work on MC in different contexts (see, e.g., [27]).

In order to capture the effect of small quantities of each chemical species in the system (i.e., S_1, S_2, S_3), we consider a stochastic model for the dynamics. To formally describe the setup, we introduce the following notation. Let $M_i^l(t)$, $l = 1, 2, 3$, $i = 1, \dots, N$ denote the random variable for the number of molecules of species S_1, S_2 or S_3 in voxel i at time t . Denote $\mathbf{M}_i(t) = [M_i^1(t), M_i^2(t), M_i^3(t)]$ as the state vector in voxel i and the matrix consisting of all state vectors by $\mathbf{M}(t) = [\mathbf{M}_1(t), \dots, \mathbf{M}_N(t)]$. The probability that $\mathbf{M}(t)$ has value \mathbf{m} at time t is then denoted by

$$P(\mathbf{m}, t) = \Pr(\mathbf{M}(t) = \mathbf{m} | \mathbf{M}(0) = \mathbf{m}_0), \quad (3)$$

where $\mathbf{M}(0)$ is the initial quantity of molecules of each species in each voxel. In the present context, the $\mathbf{M}(0)$ is dependent on the equilibrium state of the biochemical process under observation.

Since each reaction is unimolecular, it follows that in each reaction the number of molecules of the three species involved

can only increase or decrease by one. Let $\mathbf{1}_i^l$ be the state where the number of molecules in all voxels is zero, except for species l in voxel i . That is, $\mathbf{M}(t) + \mathbf{1}_i^l$ means that the number of molecules of species l in voxel i is increased by one.

A popular model for stochastic dynamics of molecules is the reaction-diffusion master equation (RDME) [28], also utilized in the context of MC in [29]. In this model, the diffusive jump rate is denoted by κ_{ij}^l for each individual molecules of the l -th species moving from voxel j into voxel i , with $\kappa_{ii} = 0$, $i = 1, \dots, N$. In particular, the probability per unit time that a molecule of S_l diffuses from voxel j to voxel i at time t is given by $\kappa_{ij}^l M_j^l(t)$. We expect that in many microfluidic systems, κ_{ij}^l is constant for a given species S_l over all voxels i, j . Nevertheless, it is also possible to consider spatially inhomogeneous diffusion [20].

Let the reactions in (1) and (2) be indexed by $r = 1, 2, 3, 4$, respectively. In the case of mass-action kinetics and first-order reactions, the probability per unit time that the substrate of reaction r , corresponding to $l = l_r$, in voxel i reacts at time t is given by $a_i^r M_i^{l_r}(t)$ with rate constants a_i^r . In general, the reaction rate is dependent on the voxel index. The net change of each chemical species due to the reaction r is expressed via the column vector $\boldsymbol{\nu}_r \in \mathbb{N}^3$. The term $\boldsymbol{\nu}_r \mathbf{1}_i$ indicates that $\mathbf{M}(t)$ changes by $\boldsymbol{\nu}_r$ in the i -th voxel. In order to model production of S_1, S_2, S_3 in the transmitter and receiver, we assume that for the voxels i comprising the transmitter $a_i^r = a^r$, $r = 1, 2$ and the receiver $a_i^r = a^r$, $r = 3, 4$, while $a_i^r = 0$ for voxels comprising the channel.

In the RMDE model, the probability distribution $P(\mathbf{m}, t)$ evolves according to the system of differential equations given by

$$\begin{aligned} \frac{dP(\mathbf{m}, t)}{dt} &= \sum_{i=1}^N \sum_{j=1}^N \sum_{l=1}^3 (\kappa_{ij}^l (m_j^l + 1) P(\mathbf{m} + \mathbf{1}_j^l - \mathbf{1}_i^l, t) \\ &\quad - \kappa_{ji}^l m_i^l P(\mathbf{m}, t)) + \sum_{i=1}^N \sum_{r=1}^4 \left(a_i^r (m_i^{l_r} + 1) P(\mathbf{m} - \boldsymbol{\nu}_r \mathbf{1}_i, t) \right. \\ &\quad \left. - a_i^r m_i^{l_r} P(\mathbf{m}, t) \right). \end{aligned} \quad (4)$$

The sums in (4) correspond to net changes in probability per unit time due to diffusion and reactions, respectively. We refer the reader to [30] for further details.

The system of ordinary differential equations in (4) corresponds to the Kolmogorov forward equation for a continuous-time Markov chain; that is, the evolution of the system state is Markovian. In our setting, the Markov chain corresponding to the RDME is irreducible and positive recurrent. Therefore, a stationary distribution exists and is given by [31]

$$\pi(\mathbf{m}) = \lim_{t \rightarrow \infty} \Pr(\mathbf{M}(t) = \mathbf{m} | \mathbf{M}(0) = \mathbf{m}_0). \quad (5)$$

III. MOLECULAR COMMUNICATION SCHEME

A. Signaling

We are concerned with the scenario that the system is reset for each experiment; i.e., when the biochemical process is ex-

posed to new environmental conditions, such as concentrations of input molecules or temperatures. In particular, no molecules of species S_1, S_2 or S_3 are present in the channel nor the spectroscopy chamber at time $t = 0$.

After an external stimulus is introduced, we assume that the biochemical process converges rapidly to an equilibrium state indexed by an element of $\{1, \dots, K\}$. For example, in the case of a unique equilibrium state (e.g., if the process is governed by a deficiency zero chemical reaction network [32]), $K = 1$. On the other hand, if the biochemical process is bistable (e.g., in quorum sensing [2]), $K = 2$. Larger numbers of equilibrium states can occur if the underlying chemical reaction network has more complex structure [22], as will be considered in Sec. V.

An equilibrium state corresponds to a concentration of S_1 denoted by C_i , $i = 1, \dots, K$. As such, the quantity of S_1 in the entrance to the microchannel is given by

$$\Delta^i = C_i V_{Tx}. \quad (6)$$

The index of the equilibrium state is denoted by $i \in \{1, \dots, K\}$.

At a time δ shortly after the biochemical process reaches equilibrium, a quantity of S_1 is present in the beginning of the microchannel, depending on equilibrium state of the biochemical process. In particular, assuming that the equilibrium state of the biochemical process is indexed by i ,

$$N_{Tx,1}(\delta) = N_{Tx,1} + \Delta^i, \quad (7)$$

where Δ^i corresponds to the quantity of molecules of S_1 in the beginning of the microchannel, as defined in (6). We note that $\delta > 0$ is chosen to be a sufficiently small period of time such that no reactions occur in the beginning of the microchannel.

The key idea behind our approach is that for sufficiently large T_s , the total number of molecules of species S_2 and S_3 in the detection chamber at the time of sampling will be approximately drawn from the stationary distribution of the RDME. As such, if the stationary distribution is known, then reliable detection of the equilibrium state for the biochemical process can be obtained.

B. Statistics for the Quantity of S_3

We now seek to obtain a good approximation for the statistics of the quantity of S_3 ; i.e., $N_3(T_s)$. Suppose that the measurement process results in the production of Δ^i at the transmitter. Then, by the analytical and empirical evaluation in [20], the following assertion provides an accurate characterization for sufficiently large T_s .

Assertion 1: Let $N_3(T_s|i)$ denote the number of molecules of S_3 in the system at time T_s given an equilibrium state of the biochemical system i corresponding to a measurement of Δ^i defined in (6). Then,

$$N_3(T_s|i) \sim \mathcal{N}(\mu^i, \mu^i), \quad (8)$$

where $\mu^i > 0$ is a known constant, only dependent on the volume of the enclosing container and not the specific geometry, and $\mathcal{N}(\mu, \sigma^2)$ denotes the Gaussian law with mean μ and variance σ^2 . In particular,

$$\mu^i = \frac{a^1}{a^2} \frac{\Delta^i \frac{V_{Rx}}{NV_{vox}}}{1 + \frac{a^1}{a^2} \frac{V_{Tx} + V_{Rx}}{NV_{vox}}}. \quad (9)$$

Under the assumption that $a^1 = a^2$ and $V_{Tx} = V_{Rx}$,

$$\mu^i = \frac{\Delta^i V_{Rx}}{NV_{vox} + 2V_{Rx}}. \quad (10)$$

So far, we have assumed that T_s is sufficiently large such that the steady state of the RDME is approximately reached. However, it is also desirable to obtain multiple observations per symbol in order to improve the detection performance. In this case, the time interval between samples is smaller and independence cannot be guaranteed from mixing properties of the RDME Markov chain. Nevertheless the convergence is rapid, as illustrated in Fig. 2, which plots the autocorrelation, $R_{ss}(m)$ of $N_{Tx,3}$ for delays m .

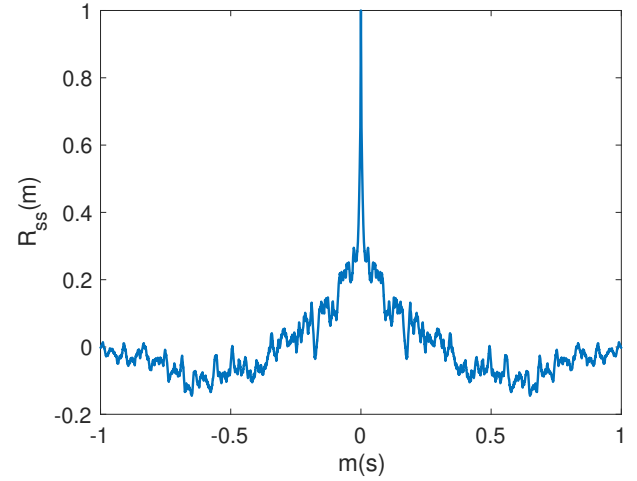


Fig. 2: Autocorrelation for $N_{Tx,3}$ with $a^1 = a^2 = 1s^{-1}$; $D_1 = D_2 = D_3 = 1.51 \times 10^{-12} \text{ m}^2/\text{s}$ and $V_{Rx} = V_{Tx}$ and $L = 40 \times 10^{-5} \text{ m}$.

Let N_s be the maximum number of independent samples that the receiver can observe per measurement; e.g., due to time constraints. Under the assumption of independent observations and the Gaussian law in Assertion 1, it is possible to reduce the variance by averaging these observations via

$$N_{Rx,3} = \frac{1}{N_s} \sum_{j=1}^{N_s} N_3(t_j|i) \quad (11)$$

where $t_j \in \{t_1, t_2, \dots, t_{N_s}\}$, where $t_j > T_s$. Then the corresponding distribution of the averaged observations $N_{Rx,3}(T_s|i)$ can be written as

$$N_{Rx,3}(T_s|i) \sim \mathcal{N}\left(\mu^i, \frac{\mu^i}{N_s}\right). \quad (12)$$

IV. BIOLOGICAL ASSAYS VIA EQUILIBRIUM SIGNALING

The aim of this work is to establish a communication mechanism to detect the equilibrium state of the biochemical process from a distributed detection chamber. While the problem bears similarities with standard formulations of MC systems—namely, that an element of a discrete set of states (or messages) is to be detected—there are two key differences.

One difference is that each equilibrium state is not in general equally likely to arise. This is due to the fact that the equilibrium state depends on the environmental input of the biochemical process and different equilibrium states are associated to a larger set of the inputs. The consequence of this is that the average probability of error is not an appropriate performance metric. Instead, the error for each equilibrium state is a more useful metric.

Perhaps the most significant difference is that the set of equilibrium states is not necessarily known before the experiment begins. This may be due to limited pre-trials, which were only able to identify a subset of the equilibrium states of the biochemical process. As such, there are three scenarios:

- (i) All equilibrium states of the biochemical process have been characterized;
- (ii) No equilibrium states of the biochemical process have been characterized;
- (iii) A subset of the equilibrium states of the biochemical process have been characterized.

For instance, initial experimentation may have revealed that various perturbations of the biochemical process lead to two distinct equilibrium states. If it is reasonable to assume there are only two states, then Scenario (i) is relevant. On the other hand, if there is some uncertainty and a strong possibility that additional equilibrium states may arise, then Scenario (iii) is relevant. If very limited initial experimentation has been performed, then Scenario (ii) is relevant.

Each scenario lends itself to a different detection or estimation strategy. In particular, Scenario (i) closely resembles a standard MC setup, where the goal is to detect an element from a known set transmitted messages. On the other hand, a desirable strategy for Scenario (ii) is to directly estimate the concentration inducing the observed signal in the receiver. Scenario (iii) lies in between Scenarios (i) and (ii).

In this section, we develop detection strategies for each of the three scenarios. In the following section (Sec. V), we evaluate their performance for biochemical processes modeled by chemical reaction systems arising in real biological systems.

A. Equilibrium State Estimation in Scenario (i)

In Scenario (i), all possible equilibrium states are known, lying in a set $\{\Delta^i\}$. By Assertion 1, the statistics for the number of molecules of S_3 in the detection chamber are known for each input of Δ^i . As such, near optimal detection can be obtained by exploiting the statistics in (12).

In particular, the optimal detection rule in the maximum likelihood sense is given by

$$\hat{s}^* = \arg \max_{i \in \{1, \dots, K\}} f(N_{R_x, 3}(T_s | i)), \quad (13)$$

where the density f is defined in (12). For most biochemical processes $K \leq 4$ and as such, evaluating the maximum likelihood detection rule is not computationally prohibitive¹.

B. Equilibrium State Estimation in Scenario (ii)

In Scenario (ii), none of the equilibrium states are known. Nevertheless, by Assertion 1, the concentration of the biochemical system can be estimated. We seek to both identify the set of equilibrium states and also estimate the equilibrium states arising from each measurement. To do so, we model the observations from a number of experiments (corresponding to changes in the environment) as a Gaussian mixture [34]. This is motivated by the fact that the observation in the detection chamber is Gaussian conditioned on the equilibrium state, with statistics given in (12).

Recall that a Gaussian mixture model is defined by a probability density function of the form

$$f_{GM}(x; \pi, \mu, \sigma) = \sum_{i=1}^N \pi_i g(x; \mu_i, \sigma_i^2), \quad (14)$$

where

$$g(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right). \quad (15)$$

In the Gaussian mixture model, the parameters μ correspond to the estimated values of the equilibrium states, σ corresponds to the standard deviation associated to each cluster, and π corresponds to the weight associated to each cluster, which is related to how likely it is that a point belongs to each cluster.

To estimate the Gaussian mixture model parameters, the standard method is based on the expectation-maximization algorithm. Suppose D measurements—corresponding to different perturbations of the biochemical process, distinct from the number of samples N_s —are taken. Then under the assumption there are K clusters, the inner loop in Algorithm 1 details the expectation-maximization procedure (i.e., the E-step and the M-step).

The parameters $\hat{\gamma}_j^i$ estimate the likelihood that measurement i is associated to cluster j , which provides the basis for clustering the observations. In particular, measurement i is associated to the cluster $j = \arg \max_k \hat{\gamma}_k^i$.

However, when the biochemical process under observation is not well understood, it may not be clear how many equilibrium points and hence which cluster size K should be chosen. In order to select the cluster size K , we exploit the Akaike information criterion (AIC) [35], detailed in line 9 of Algorithm 1. In particular, AIC is given by

$$AIC = 2K - 2L \quad (16)$$

where L represents the maximized log likelihood function given in line 8 of Algorithm 1. As detailed in Algorithm 1, the GMM is applied for different values of K (corresponding to a different number of equilibrium states). The value of K then corresponds to the minimum AIC.

¹We note that Scenario (i) can also be viewed as a form of multi-level equilibrium signaling developed further in [33].

The AIC can be replaced by the Bayesian Information Criterion (BIC) [36], given by

$$BIC = K \log(D) - 2L. \quad (17)$$

This can be implemented in Algorithm 1 by replacing the AIC in line 8 with the BIC.

Algorithm 1 GMM Clustering Algorithm for Scenario (ii)

- 1: For $K = 1 : K_{max}$
 - 2: **Initialize:** $\hat{\mu}_{1:K}, \hat{\sigma}_{1:K}^2, \pi_{1:K}$
 - 3: While not converged
 - 4: $\phi_j \sim \mathcal{N}(\hat{\mu}_j, \hat{\sigma}_j^2) \quad j = 1 : K$
 - 5: E-step : $\hat{\gamma}_j^i = \frac{\pi_j \phi_j(x_i)}{\sum_{k=1}^K \pi_k \phi_k(x_i)} \quad j = 1 : K, \quad i = 1 : D$
 - 6: M-step : $\hat{\mu}_j = \frac{\sum_{i=1}^D \hat{\gamma}_j^i x_i}{\sum_{i=1}^D \hat{\gamma}_j^i}$
 $\hat{\sigma}_j^2 = \frac{\sum_{i=1}^D \hat{\gamma}_j^i (x_i - \hat{\mu}_j)^2}{\sum_{i=1}^D \hat{\gamma}_j^i} \quad j = 1 : K$
 $\pi_j = \frac{\sum_{i=1}^D \hat{\gamma}_j^i}{D}$
 - 7: End while
 - 8: $L = \sum_{i=1}^D \log \sum_{j=1}^K \pi_j \phi_j(x_i)$
 - 9: $AIC_K = 2K - 2L$
 - 10: End for
 - 11: Return $K_{opt} = \min_K AIC_K, \quad \hat{\mu}_{1:K_{opt}}$
-

C. Equilibrium State Estimation in Scenario (iii)

In Scenario (iii), only a subset of the equilibrium states $\{\Delta^i\}$ are known. The goal is therefore to identify whether each observation corresponds to a known equilibrium state, and if not, estimate the new equilibrium state.

In order to estimate the equilibrium states, we follow a similar approach to Algorithm 1 for Scenario (ii). The key difference, as detailed in Algorithm 2, is that since some equilibrium states are known, the means of some clusters can be initialized as these equilibrium states. Then, during the expectation-maximization algorithm, the means of these clusters are not updated. Moreover, in applying the AIC, the minimum number of clusters is given by the number of known equilibrium states. As for Algorithm 1, line 10 in Algorithm 2 can be replaced with the BIC defined in (17).

V. EVALUATION IN REALISTIC BIOCHEMICAL PROCESSES

A. Self Activating Toggle Switch Circuits

Gene regulatory networks play a key role in cellular differentiation. In particular, the decision networks often heavily rely on a regulatory motif in which the proteins of two genes act as mutually inhibiting transcription factors [22]. It is often the case that one or both genes also act as self-activating transcription factors, leading to a self-activating toggle switch. A key example is the (CDX2-OCT3/4) pair [22], which is associated with the differentiation of embryonic stem cells.

An important feature of self-activating toggle switches is that they can admit tristability. That is, three distinct stable equilibrium states are present. Moreover, the occurrence of distinct equilibrium states varies from one equilibrium state

Algorithm 2 GMM Clustering Algorithm for Scenario (iii)

- 1: Input: Known K_u Equilibrium points and corresponding means $\mu_{1:K_u}$
 - 2: For $K = K_u : K_{max}$
 - 3: **Initialize:** $\hat{\mu}_{1:K}, \hat{\sigma}_{1:K}^2, \pi_{1:K}$
 - 4: While not converged
 - 5: $\phi_j \sim \mathcal{N}(\hat{\mu}_j, \hat{\sigma}_j^2) \quad j = 1 : K$
 - 6: E-step : $\hat{\gamma}_j^i = \frac{\pi_j \phi_j(x_i)}{\sum_{k=1}^K \pi_k \phi_k(x_i)} \quad j = 1 : K, \quad i = 1 : D$
 - 7: M-step : $\hat{\mu}_j = \mu_j, \hat{\sigma}_j^2 = \frac{\mu_j^2}{N_s} \quad j = 1 : K_u$
 $\hat{\mu}_j = \frac{\sum_{i=1}^D \hat{\gamma}_j^i x_i}{\sum_{i=1}^D \hat{\gamma}_j^i} \quad j = K_u + 1 : K$
 $\hat{\sigma}_j^2 = \frac{\sum_{i=1}^D \hat{\gamma}_j^i (x_i - \hat{\mu}_j)^2}{\sum_{i=1}^D \hat{\gamma}_j^i} \quad j = 1 : K$
 $\pi_j = \frac{\sum_{i=1}^D \hat{\gamma}_j^i}{D}$
 - 8: End while
 - 9: $L = \sum_{i=1}^D \log \sum_{j=1}^K \pi_j \phi_j(x_i)$
 - 10: $AIC_K = 2K - 2L$
 - 11: End for
 - 12: Return $K_{opt} = \min_K AIC_K, \quad \hat{\mu}_{1:K_{opt}}$
-

to another. As a consequence, some equilibrium states can be much rarer than others.

We consider an experiment where the goal is to identify the equilibrium states corresponding to different initial conditions of a self-activating toggle switch, when none, some, or all equilibrium states are known. In particular, the system consists of two transcription factors and two noncompetitive binding sites, which facilitate transcription of proteins A and B.

The self-activating toggle switch is governed by the system of ordinary differential equations [22]

$$\begin{aligned} \frac{dc_A}{dt} &= G_A(c_A, c_B) - k_A c_A \\ \frac{dc_B}{dt} &= G_B(c_B, c_A) - k_B c_B, \end{aligned} \quad (18)$$

where the mean transcription rate of X is given by

$$\begin{aligned} G_X(c_A, c_B) &= (g_{X,AB} + g_{X,A} H^-(c_B)) H^+(c_A) + g_{X,B} + g_{X,0} H^-(c_B), \end{aligned} \quad (19)$$

with transcription rates for each promoter state given by $g_{X,A} \gg g_{X,0}, g_{X,B} \ll g_{X,0}, g_{X,AB} \approx g_{X,0}$, and the dynamics are governed by the Hill functions

$$\begin{aligned} H^-(c_A) &= \frac{1}{1 + c_A^{n_A}} \\ H^+(c_A) &= 1 - H^-(c_A). \end{aligned} \quad (20)$$

Fig. 3 plots the equilibrium states arising from the model in (18) with parameters detailed in the figure. In this figure, equilibrium concentration of species A (denoted by c_A) is plotted for different initial concentration of species A and species B. Observe that the system is tristable, with three equilibrium states that occur with different frequencies. Moreover, each equilibrium state does not occur with the same frequency. In other words, as can be seen in Fig. 3, the frequency of an equilibrium state indicated by blue is much smaller than the states indicated by cyan and magenta.

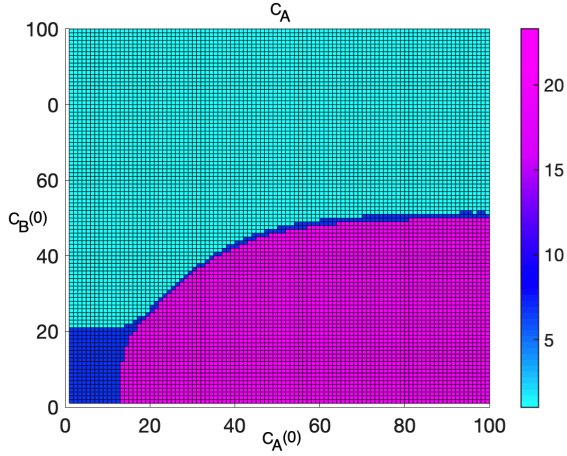


Fig. 3: Equilibrium states of type A for different initial inputs of A and B in the toggle switch system with $n_A = 4$, $k_A = 4s^{-1}$, $k_B = 4s^{-1}$, $g_{A,A} = 5s^{-1}$, $g_{A,0} = 1s^{-1}$, $g_{A,B} = 0.1s^{-1}$, $g_{A,AB} = g_{A,0}$, $g_{B,A} = 0.1s^{-1}$, $g_{B,0} = 1s^{-1}$, $g_{B,B} = 10s^{-1}$, $g_{B,BA} = g_{B,0}$ resulting in three different equilibrium points $\mu^1 = 1.01$, $\mu^2 = 6.64$, $\mu^3 = 23.29$.

We now evaluate the performance of each algorithm in Section IV for the biochemical process defined by (18) with parameters the same as in Fig. 3. In particular, Fig. 4 shows the probability that an observation is incorrectly clustered. In the case of Scenario (i), this means that the wrong equilibrium state has been identified and can be interpreted as the clustering error. Observe that as the number of samples for each measurement increases, the clustering error decreases. This is expected as increasing the number of samples reduces the variance of the approximate Gaussian model in (12). We also observe that the clustering error differs for each of the three equilibrium points. This is also expected since different equilibrium points have different variances at the detection chamber, which follows from (12).

In the case of Scenario (ii), no equilibrium states are initially known to the receiver. In order to estimate the equilibrium states and identify which equilibrium state each observation should be assigned to, we utilize Algorithm 1. In this case, the clustering error is the proportion of points that are not associated with the cluster with mean closest to the true equilibrium state for the point. Observe that again, the clustering error decreases with an increase in the number of samples, as for Scenario (i).

In Scenario (iii), the second and the third equilibrium points are assumed to be known. While the first equilibrium point and the total number of actual equilibrium points are initially not known. The estimates of the unknown equilibrium states are found via Algorithm 2. As for Scenario (i) and Scenario (ii), the clustering error decreases for an increasing number of samples. We also observe that the clustering error is slightly higher than the Scenario (ii) and slightly lower than Scenario (i). This is also expected since in Scenario (i), the set of equilibrium states is completely known, while in Scenario (iii) there is only partial knowledge of the set of equilibrium states.

In Fig. 5, the estimation error of the equilibrium states in

Scenario (ii) and Scenario (iii) is plotted. The parameters for the tristable model in (18) are the same as for Fig. 4. For Scenario (ii), corresponding to Fig. 5a, no equilibrium states are known. Observe that the estimation error for all states as the number of samples increases. As expected, the estimation error measured by the normalized mean-square error (NMSE) decreases as the number of samples increases.

Similar behavior is exhibited in Fig. 5b for Scenario (ii) where only the third equilibrium state is unknown. Indeed, the NMSE again decreases as the number of samples is increased. Moreover, the estimation error is significantly smaller than for Scenario (ii). This can be explained by the fact that a fewer number of parameters are required to be estimated via the Gaussian mixture model.

B. Robust Adaptation

In order for biological systems to adapt to changes in their environment, a means of sensing is required in order to determine the concentration of various chemical species in their environment. This kind of sensing mechanism, known as *chemosensing*, involves two key aspects: receptors on the surface of the biological system to bind with chemicals in the environment; and a network of chemical reactions to produce a desired response *within* the biological system. This response can be quantified in terms of the steady-state concentration levels of certain chemical species within the biological system.

A popular example of chemosensing arises in bacteria chemotaxis in *E. coli* [23], where each bacterium seeks to move closer to attractants such as glucose and away from repellants such as phenol. This is achieved via proteins on the surface of the bacterium's membrane, which bind to attractants and repellants resulting in changes to steady state concentrations of proteins within the bacterium (e.g., phospho-CheY). The consequence of this change in concentration of phospho-CheY is an increase or decrease in the probability a rotor attached to each bacterium spins. This in turn leads to changes in the direction of the bacterium.

An important feature of the chemosensing mechanism in *E. coli* is that the phospho-CheY response should remain approximately constant when the concentration of attractants or repellants exceeds a given level, known as *robust adaptation*. This is to ensure that the rotor control mechanism is not overwhelmed.

We now consider the scenario where a biological assay is performed in order to study robust adaptation, such as in bacterial chemotaxis. We assume that the biological system is governed by the chemical reaction network given by [23]



Let $c_A(t)$ and $c_B(t)$ denote the concentration of species A and B, respectively. Define $\Theta = c_A(0) + c_B(0)$. Then, the unique *positive* equilibrium concentrations under mass action

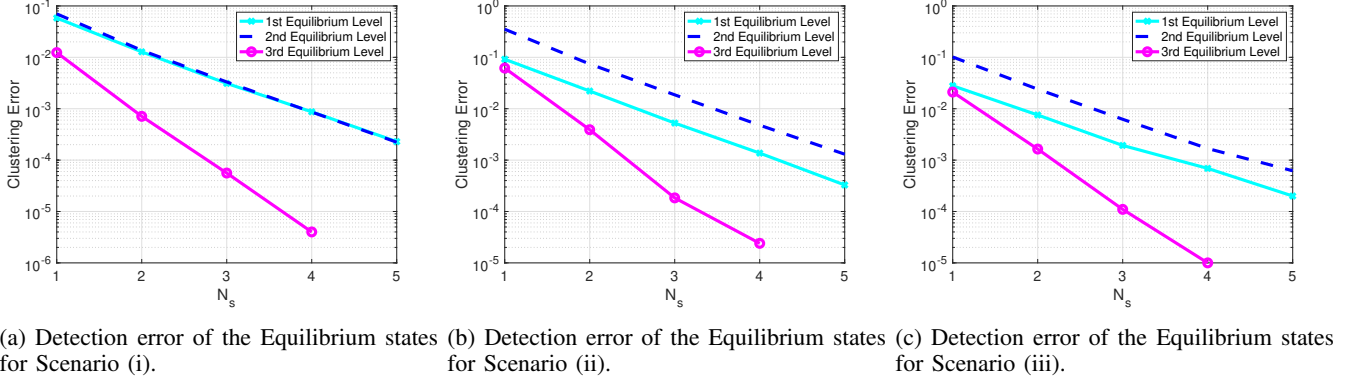


Fig. 4: Detection error of the equilibrium states for different scenarios in Self Activating Toggle Switch Circuits

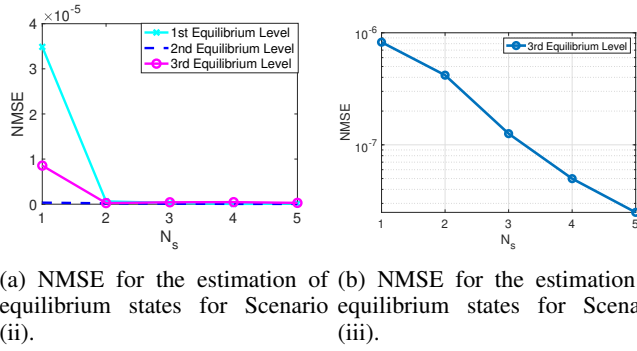


Fig. 5: Estimation error of the equilibrium states in the self activating toggle switch model.

kinetics are given by

$$\begin{aligned} c_A^\infty &= \frac{\beta}{\alpha} \\ c_B^\infty &= \Theta - \frac{\beta}{\alpha}. \end{aligned} \quad (22)$$

Suppose from a preliminary investigation, it is suspected that there is a unique positive equilibrium corresponding to $c_A^\infty = \frac{\beta}{\alpha}$. The purpose of the assay is to verify that there are no larger positive equilibria.

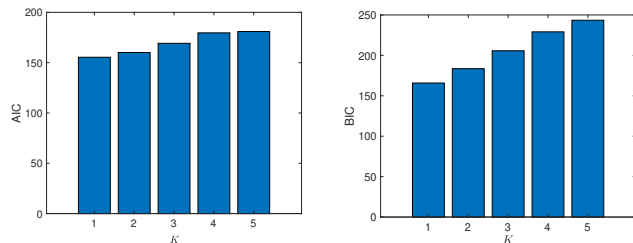


Fig. 6: Selection of K for the robust adaptation model using AIC and BIC.

In order to do so, we apply Algorithm 2 corresponding to Scenario (iii), where a subset of equilibrium states are known

for $\beta = 0.5$ and $\alpha = 0.1$. Since the system has only one positive equilibrium state, we should expect that the Gaussian mixture model has only a single cluster. That is, the AIC and BIC is minimized for a single cluster. To confirm this hypothesis, Fig. 6 plots the AIC and the BIC for varying choices of K , corresponding to the number of equilibrium states. Observe that the AIC and BIC are minimized for a single cluster, consistent with the model.

VI. CONCLUSION

Advanced spectroscopy methods provide a means of passively estimating the number of molecules of a given chemical species. One application is to investigate the equilibrium behavior of a range of biochemical processes. However, spectroscopy devices may require special conditions to operate, which are not consistent with normal behavior of the biochemical process. In this case, it is desirable to exploit LoC technology to measure chemical signals in a location away from where the biochemical process.

In order to do so, it is necessary to introduce a communication channel and a means of characterizing the statistics for the number of molecules in the external spectroscopy chamber. In this paper, we have proposed an approach exploiting the steady state behavior of reaction-diffusion systems. We have shown that this approach yields reliable estimates of equilibrium states, even if the full set of possible equilibrium states is not known.

In particular, we have studied the tradeoff between estimation error in the equilibrium states with the measurement rate required for spectroscopy. This provides a basis for further work investigating spectroscopy methods in the context of LoC devices for high-throughput biological assays.

REFERENCES

- [1] U. Alon, M. Surette, N. Barkai, and S. Leibler, "Robustness in bacterial chemotaxis," *Nature*, vol. 397, no. 6715, pp. 168–171, 1999.
- [2] M. Miller and B. Bassler, "Quorum sensing in bacteria," *Annual Reviews in Microbiology*, vol. 55, no. 1, pp. 165–199, 2001.
- [3] F. Freyer, J. Roberts, P. Ritter, and M. Breakspear, "A canonical model of multistability and scale-invariance in biological systems," *PLoS Computational Biology*, vol. 8, no. 8, 2012.
- [4] W. Pearman, M. Lawrence-Snyder, S. Angel, and A. Decho, "Surface-enhanced Raman spectroscopy for *in situ* measurements of signaling molecules (autoinducers) relevant to bacteria quorum sensing," *Applied Spectroscopy*, vol. 61, no. 12, pp. 1295–1300, 2007.

- [5] J. Zlatanova *et al.*, “Single molecule force spectroscopy in biology using the atomic force microscope,” *Progress in Biophysics and Molecular Biology*, vol. 74, no. 1-2, pp. 37–61, 2000.
- [6] L. Ferrari, J. Kaufmann, F. Winnefeld, and J. Plank, “Interaction of cement model systems with superplasticizers investigated by atomic force microscopy, zeta potential, and adsorption measures,” *Journal of Colloid and Interface Science*, vol. 347, pp. 15–24, 2010.
- [7] R. Fair, “Digital microfluidics: is a true lab-on-a-chip possible?” *Microfluidics and Nanofluidics*, vol. 3, no. 3, pp. 245–281, 2007.
- [8] M. Kusec and O. Akan, “Modeling convection-diffusion-reaction systems for microfluidic molecular communications with surface-based receivers in internet of bio-nano things,” *PLoS One*, vol. 13, no. 2, 2018.
- [9] Y. Deng, M. Pierobon, and A. Nallanathan, “A microfluidic feed forward loop pulse generator for molecular communication,” in *IEEE Global Communications Conference (GLOBECOM)*, 2017.
- [10] D. Bi, Y. Deng, M. Pierobon, and A. Nallanathan, “Chemical reactions-based microfluidic transmitter and receiver for molecular communication,” *arXiv:1908.03441v1*, 2019.
- [11] E. De Leo, L. Donvito, L. Galluccio, A. Lombardo, G. Morabito, and L. Zanolli, “Communications and switching in microfluidic systems: pure hydrodynamic control for networking labs-on-a-chip,” *IEEE Transactions on Communications*, vol. 61, no. 11, pp. 4663–4677, November 2013.
- [12] M. Hamidović *et al.*, “Information encoding in droplet-based microfluidic systems: first practical study,” in *Proc. ACM International Conference on Nanoscale Computing and Communication*, 2019.
- [13] N. Farsad, A. Eckford, S. Hiyama, and Y. Moritani, “On-chip molecular communication: analysis and design,” *IEEE Transactions on NanoBioscience*, vol. 11, no. 3, pp. 304–314, September 2012.
- [14] A. O. Bicen, J. J. Lehtomäki, and I. F. Akyildiz, “Shannon meets Fick on the microfluidic channel: diffusion limit to sum broadcast capacity for molecular communication,” *IEEE Transactions on NanoBioscience*, vol. 17, no. 1, pp. 88–94, January 2018.
- [15] B. Krishnaswamy *et al.*, “Time-elapse communication: bacterial communication on a microfluidic chip,” *IEEE Transactions on Communications*, vol. 61, no. 12, pp. 5139–5151, December 2013.
- [16] L. Galluccio *et al.*, “Capacity of a binary droplet-based microfluidic channel with memory and anticipation for flow-induced molecular communications,” *IEEE Transactions on Communications*, vol. 66, no. 1, pp. 194–208, January 2018.
- [17] A. Bicen and I. Akyildiz, “System-theoretic analysis and least-squares design of microfluidic channels for flow-induced molecular communication,” *IEEE Transactions on Signal Processing*, vol. 61, no. 20, pp. 5000–5013, October 2013.
- [18] R. Van Dam, “Solvent-resistant elastomeric microfluidic devices and applications,” *Diss. California Institute of Technology*, 2006.
- [19] E. Seitaridou *et al.*, “Measuring flux distributions for diffusion in the small-numbers limit,” *The Journal of Physical Chemistry B*, vol. 111, no. 9, pp. 2288–2292, 2007.
- [20] B. Akdeniz, M. Egan, and B. Tang, “Equilibrium signaling: molecular communication robust to geometry uncertainties,” *IEEE Transactions on Communications*, 2020.
- [21] H. Shahmohammadian, G. Messier, and S. Magierowski, “Optimum receiver for molecule shift keying modulation in diffusion-based molecular communication channels,” *Nano Communication Networks*, vol. 3, no. 3, pp. 183–195, 2012.
- [22] M. Lu *et al.*, “Tristability in cancer-associated microRNA-TF chimera toggle switch,” *J. Phys. Chem. B*, vol. 117, pp. 13 164–13 174, 2013.
- [23] G. Shinar and M. Feinberg, “Structural sources of robustness in biochemical reaction networks,” *Science*, vol. 327, pp. 1389–1391, 2010.
- [24] L. Zeman and A. Zydney, *Microfiltration and Ultrafiltration: Principles and Applications*. CRC Press, 2017.
- [25] H. Guo, J. M. Jornet, Q. Gan, and Z. Sun, “Cooperative Raman spectroscopy for real-time in vivo nano-biosensing,” *IEEE Transactions on NanoBioscience*, vol. 16, no. 7, pp. 571–584, 2017.
- [26] D. Soloveichik, G. Seelig, and E. Winfree, “DNA as a universal substrate for chemical kinetics,” *Proceedings of the National Academy of Sciences*, vol. 107, no. 12, pp. 5393–5398, 2010.
- [27] A. Noel, K. C. Cheung, and R. Schober, “Optimal receiver design for diffusive molecular communication with flow and additive noise,” *IEEE transactions on nanobioscience*, vol. 13, no. 3, pp. 350–362, 2014.
- [28] S. Isaacson, “The reaction-diffusion master equation as an asymptotic approximation of diffusion to a small target,” *SIAM Journal on Applied Mathematics*, vol. 70, no. 1, pp. 77–111, 2009.
- [29] C. Chou, “Extended master equation models for molecular communication networks,” *IEEE Transactions on NanoBioscience*, vol. 12, no. 2, pp. 79–92, 2013.
- [30] N. van Kampen, *Stochastic Processes in Physics and Chemistry*. Elsevier Science & Technology Books, 2007.
- [31] S. Ethier and T. Kurtz, *Markov Processes: Characterization and Convergence*. John Wiley & Sons, 2009.
- [32] M. Feinberg, “Chemical reaction network structure and the stability of complex isothermal reactors—I. The deficiency zero and deficiency one theorems,” *Chemical Engineering Science*, vol. 42, no. 10, pp. 2229–2268, 1987.
- [33] B. Akdeniz and M. Egan, “Multi-level equilibrium signaling for molecular communication,” in *Proc. ACM International Conference on Nanoscale Computing and Communication*, 2020.
- [34] D. Reynolds, “Gaussian mixture models,” *Encyclopedia of Biometrics*, vol. 741, 2009.
- [35] H. Akaike, “A new look at the statistical model identification,” *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, 1974.
- [36] G. Schwarz, “Estimating the dimension of a model,” *Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978.